yw

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/031,025	06/06/2002	Albertus Cornelis Den Brinker	NL 000288	4466

7590 02/18/2005

Corporate Patent Counsel
Philips Electronics North America Corporation
580 White Plains Road
Tarrytown, NY 10591

EXAMINER	
VO, HUYEN X	

ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 02/18/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

RECEIVED
MAR 0 2 2005
Technology Center 2600

PTO 90C (Rev 10/03)

Office Action Summary	Application No. 10/031,025	Applicant(s) DEN BRINKER ET AL.	
	Examiner Huyen Vo	Art Unit 2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06 June 2002*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-7* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-7* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *06 June 2002* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *6/6/02*.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless – (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 1, 3, and 5-6 are rejected under 35 U.S.C. 102(b) as being anticipated by Eatwell (US Patent No. 5742694).

3. Regarding claims 1, 3, and 5, Eatwell discloses a method of encoding an audio signal, an audio encoder and system, comprising the steps of:

determining basic waveforms in the audio signal (*Predictable Component 3 in figure 4*);

obtaining a noise component from the audio signal by subtracting the basic waveforms from the audio signal (*Prediction Error 4 in figure 4*);

modeling a spectrum of the noise component by determining auto-regressive and moving-average parameters (*element 42 in figures 3 or 5, or referring to col. 8, ln. 22-47, where filter coefficients are determined*); and

including the auto-regressive and the moving-average parameters

and waveform parameters representing the basic waveforms in an encoded audio signal (*Output signal 8 in figures 3 or 5*).

4. Regarding claim 6, Eatwell discloses an encoded audio signal comprising:

waveform parameters representing basic waveforms (*Predictable Component 3 in figure 4*); and

auto-regressive parameters and moving-average parameters representing a spectrum of a remaining noise component (*element 42 in figures 3 or 5 or referring to col. 8, ln. 22-47*).

5. Claims 1-6 are rejected under 35 U.S.C. 102(e) as being anticipated by Miseki et al. (US Patent No. 6167375).

6. Regarding claims 1, 3, and 5, Miseki et al. disclose a method of encoding an audio signal, an audio encoder and system, comprising the steps of:

determining basic waveforms in the audio signal (*Predictor 547 in figure 18*);

obtaining a noise component from the audio signal by subtracting the basic waveforms from the audio signal (*Output of the summer 543 in figure 18*);

modeling a spectrum of the noise component by determining auto-regressive and moving-average parameters (*col. 8, ln. 22-47, where filter coefficients are determined*); and

including the auto-regressive and the moving-average parameters

and waveform parameters representing the basic waveforms in an encoded audio signal (*Output signal 513 in figure 16*).

7. Regarding claim 6, Miseki et al. discloses an encoded audio signal comprising:

waveform parameters representing basic waveforms (*Predictor 547 in figure 18*); and auto-regressive parameters and moving-average parameters representing a spectrum of a remaining noise component (*col. 8, ln. 22-47, filter coefficients are determined*).

8. Regarding claims 2, 4, and 5, Miseki et al. disclose a method of decoding an encoded audio signal, an audio player and system, comprising the steps of:

receiving an encoded audio signal comprising waveform parameters representing basic waveforms and auto-regressive and moving-average parameters representing a spectrum of a remaining noise component (*col. 23, ln. 1-35 together with figure 20, ARMA parameters are transmitted to the decoder side for use*);

filtering a white noise signal to obtain a reconstructed noise component, which filtering is determined by the auto-regressive parameters and the moving-average parameters (*Noise Decoder 290 in figure 23*);

synthesizing basic waveforms based on the waveform parameters (*Speech Decoder 280 in figure 23*); and

adding the reconstructed noise component to the synthesized basic waveforms to obtain a decoded audio signal (*Mixer 295 in figure 23*).

Claim Rejections - 35 USC § 103

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Eatwell (US Patent No. 5742694).

11. Regarding claim 7, Eatwell fails to disclose a storage medium on which an encoded audio signal as claimed in claim 6 is stored. However, it would have been obvious to one skilled in the art at the time of invention to implement the method in claim 6 in computer codes to facilitate maintenance and updating.

12. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Miseki et al. (US Patent No. 6167375).

13. Regarding claim 7, Miseki et al. fail to disclose a storage medium on which an encoded audio signal as claimed in claim 6 is stored. However, it would have been obvious to one skilled in the art at the time of invention to implement the method in claim 6 in computer codes to facilitate maintenance and updating.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. *Akamine* et al. (IEEE Publication) teach an ARMA model based speech coding scheme that is considered pertinent to the claimed invention.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Huyen Vo whose telephone number is 703-305-8665. The examiner can normally be reached on M-F, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Doris To can be reached on 703-305-4827. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Examiner Huyen X. Vo November 22, 2004

***

Susan McFadden
SUSAN MCFADDEN
PRIMARY EXAMINER

OIPE
JUN 11 2002
JC22
PATENT & TRADEMARK OFFICE

Form PTO-1449 U.S. DEPARTMENT OF COMMERCE (REV. 7-80) PATENT AND TRADEMARK OFFICE	Atty. Docket No. NL 000288	Serial No. 10/031,025
	Applicant ALBERTUS C. DEN BRINKER ET AL.	
INFORMATION DISCLOSURE CITATION (Use several sheets if necessary)	Filing Date 1/14/02	Group

U.S. PATENT DOCUMENTS

Ex. Int.		Document Number	Date	Name	Class	Sub-class	Filing Date If Approp.
	AA						
	AB						
	AC						
	AD						
	AE						
	AF						

RECEIVED
AUG 16 2002
Technology Center 2600

FOREIGN PATENT DOCUMENTS

		Document Number	Date	Country	Class	Sub-class	Trans. Yes	Trans. No
	AG							
	AH							
	AI							
	AJ							
	AK							

OTHER (Including Author, Title, Date, Pertinent Pages, Etc.)

HV	AL	Petre Stoica et al., "Introduction to Spectral Analysis," Prentice-Hall Inc., 1997, pages 101-108.
	AM	
	AN	

Examiner Huyen X Vo	Date Considered 11/22/04

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

S:\SL\NL000288DISCL.SL.DOC

531 Rec'd PCT/PTC 14 JAN 2002

Sheet 1 of 1

Form PTO-1449 U.S. DEPARTMENT OF COMMERCE (REV. 7-80) PATENT AND TRADEMARK OFFICE	Atty. Docket No. NL 000288	Serial No. 10/031025
	Applicant ALBERTUS C. DEN BRINKER ET AL.	
INFORMATION DISCLOSURE CITATION (Use several sheets if necessary)	Filing Date CONCURRENTLY	Group

U.S. PATENT DOCUMENTS

Ex. Int.		Document Number	Date	Name	Class	Sub-class	Filing Date If Approp.
HV	AA	4188667	2/12/80	GRAUPE ET AL.	364	724	
HV	AB	5943429	8/24/99	HANDEL	381	94.1	
	AC						
	AD						
	AE						
	AF						

FOREIGN PATENT DOCUMENTS

		Document Number	Date	Country	Class	Sub-class	Trans. Yes	Trans. No
HV	AG	WO9728527	07.08.97	PCT (WORLD)	G10L	3/01		
	AH							
	AI							
	AJ							
	AK							

OTHER (Including Author, Title, Date, Pertinent Pages, Etc.)

	AL	
	AM	
	AN	

Examiner Huyen X Vo   Date Considered 11/22/04

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP Draw line through citation if not in conformance and not considered. Include a copy this form with next communication to applicant.

S:\SL\NL000288DISCL.MA0.doc

Notice of References Cited	Application/Control No. 10/031,025	Applicant(s)/Patent Under Reexamination DEN BRINKER ET AL.	
	Examiner Huyen Vo	Art Unit 2655	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-5,742,694	04-1998	Eatwell, Graham P.	381/94.2
	B	US-6,167,375	12-2000	Miseki et al.	704/229
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Akamine, M.; Miseki, K., "ARMA model based speech coding at 8 kb/s" Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International, Conference on , 23-26 May 1989, Pages:148 - 151 vol.1
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

S4.8

ARMA MODEL BASED SPEECH CODING AT 8KB/S

Masami AKAMINE and Kimio MISEKI

Video Sys. & Tech. Lab., Toshiba R & D Center
Saiwai-ku, Kawasaki-shi, Japan

ABSTRACT

This paper proposes a new speech coding for high quality performance at 8 kbps. The coder is based on an ARMA model as the prediction filter and a new excitation model. A simple ARMA analysis method is proposed. This analysis method features a technique for eliminating the fine harmonic structure within the speech spectrum. This overcomes the problem of MA parameter misestimation. The excitation signal is modeled as a pulse train whose density is varied depending on the residual signal's power. The proposed coder produced high quality speech comparable with 6-bit log PCM at 8 kbps according to computer simulation.

1. INTRODUCTION

Various speech coding methods at bit rate below 10 kbps have been proposed for applications to private comunication networks and mobile systems. A class of methods, adaptive predictive coding (APC) [1][2], multi pulse coding (MPC) [3][4], and code excited linear prediction coding (CELP) [5] seems to be promising for these applications. This class of methods represents a speech spectral envelope as an all pole model, in other words, an AR model. However there are not only poles but also zeros in speech spectra. Especially, nasal and consonant sounds tend to have spectral zeros. The AR model is less satisfactory in describing these zeros accurately. An ARMA model can be attractive for improving speech quality in low bit rate speech coding because of its ability to describe both spectral poles and zeros efficiently. However, the ARMA model has rarely been used in speech coding so far, because the resulting estimation problem in an ARMA based system is a non-linear problem.

Ishizaki proposed a simple ARMA analysis method [6]. In the method, the AR and MA parameters are obtained separately. The AR parameters are determined by LPC analysis using the autocorrelation method. The MA parameters are obtained by inverting the power spectrum of the AR model's residual signal, and applying LPC analysis to the corresponding autocorrelation function. This ARMA estimation method requires LPC analysis and FFT. Therefore, the method has low computational complexity. But, the method often fails to estimate the MA parameters because of the fine harmonic structure within speech spectra. Thus, a technique for eliminating the fine harmonic structure in the frequency domain and the time domain has been considered here.

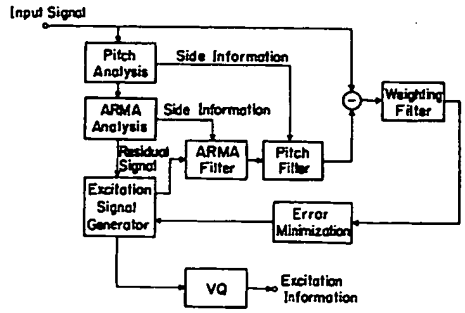

Fig. 1 Block diagram of the proposed coding algorithm.

The authors propose a new ARMA model based speech coder. The proposed coder includes a pitch filter, an ARMA filter, and an excitation signal model. The excitation signal is modeled as a pulse train whose density is varied depending on the residual signal's power. The ARMA residual signal is divided into several subframe signals in the time domain. The excitation signal is represented by closely spaced pulses in subframes with large residual power, while the excitation signal is represented by more widely spaced pulses in the subframes with low residual power. Each subframe's excitation is determined analytically to minimize the perceptually weighted errors between original and synthetic signals.

2. CODING ALGORITHM

Figure 1 shows a block diagram of the proposed coder. The coder consists of two parts. The first part includes a pitch filter and an ARMA filter. The second part is an excitation signal generator. The pitch filter generates the pitch periodicity of voiced speeches. The ARMA filter restores the spectral envelope. The excitation signal is modeled as adaptive density pulses (ADP). Its amplitude is quantized with a vector quantizer. The pitch parameters, the ARMA parameters and the ADP's density and phase are transmitted to the receiver as side information. The coder will be called ARMA-ADP in this paper.

A. ARMA ANALYSIS

The proposed analysis method is depicted in Fig.2. In this method, the AR and MA parameters are estimated separately. The well known LPC analysis method is used for estimating the AR

CH2673-2/89/0000-0148 $1.00 © 1989 IEEE

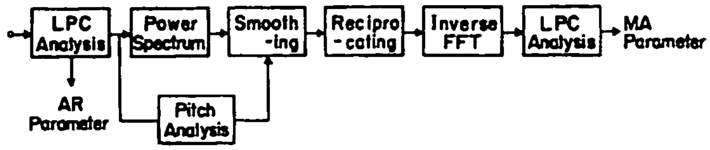

Fig. 2 Block diagram of the proposed ARMA analysis method.

parameters. After converting spectral zeros to poles, LPC analysis is applied to obtain the MA parameters. Zero-to-pole conversion is accomplished as follows. First, the power spectrum of the AR residual signal is calculated by FFT. Then the power spectrum is filtered in the frequency domain to remove its spectral harmonic structure. This filtering is equivalent to spectrum smoothing. The proposed ARMA analysis has a feature in the method for removing the harmonic structure. If we fail to remove it, the MA parameters are misestimated, because the deep valleys of the harmonic structure are changed to acute peaks in the zero-to-pole conversion.

The authors considered the ways for eliminating the fine harmonic structure. Voiced speech waveforms are quasi periodic, not strictly periodic. However, for the sake of simple explanation, voiced speech is assumed to be strictly periodic signal with a pitch period Tp[s]. Then, the speech spectra are frequency-discrete. Considering the Fourier transform of the discrete spectra, the Fourier representation is a periodic function with a period Tp. A continuous spectrum can be obtained by using an ideal lowpass filter with a Tp/2 bandwidth. Therefore, the frequency characteristic of the filter must change adaptively depending on the pitch period. Also, the filter must have a zero-phase characteristic. Spectrum zeros will be shifted unless the filter is zero-phase. It should be noticed that a filter closely approximating an ideal lowpass filter is computationally expensive because the resulting filter is of a high order. A low order filter causes less distortion in the spectral envelope because the autocorrelation function r(k) of the AR residual signal decreases rapidly with k.

Thus, in order to remove the fine structure, a first-order IIR filter was used whose coefficient is adaptively varied so that the filter's time constant is proportional to the interval of the pitch harmonics. Let N be the order of FFT and T [sample] the pitch period. The filter's coefficient a is described as a function of the pitch period as follows.

$$a = \exp(Nb/T) \qquad (1)$$

where b is a constant which is experimentally determined.
Filtering is accomplished toward two directions, forward and backward as follows.

$$D_f(k) = D(k) + a\, D_f(k-1) \qquad k=1,2,\cdots,N \qquad (2)$$

$$D_b(k) = D(k) + a\, D_b(k+1) \qquad k=N,N-1,\cdots,1 \qquad (3)$$

$$\bar{D}(k) = [D_f(k) + D_b(k)]\,/2 \qquad k=1,2,\cdots,N \qquad (4)$$

where $D_f(k)$, $D_b(k)$, $D(k)$ and $\bar{D}(k)$ are the forward filtered spectrum, the backward filtered spectrum, the power spectrum of the AR residual signal and the smoothed power spectrum, respectively. This makes the filter's phase characteristic zero-phase.

The operations described above are done in frequency domain. The power spectrum smoothing and the zero-to-pole conversion can also be carried out in the time domain. Spectrum smoothing is carried out by windowing the corresponding autocorrelation function with a rectangular window. Zero-to-pole conversion is performed as follows. Let $D_r(k)$ be the reciprocal of the smoothed power spectrum $\bar{D}(k)$. Then

$$\bar{D}(k)\cdot D_r(k) = 1 \qquad k=1,2,\cdots,N. \qquad (5)$$

By the inverse Fourier transform of the above equation, we can obtain

$$\sum_{k=0}^{N-1} r(k)\,r'(n-k) = \delta(n) \qquad n=0,1,\cdots,N-1 \qquad (6)$$

where r(k) and r'(k) is the corresponding autocorrelation function of $\bar{D}(k)$ and Dr(k), respectively, and $\delta(n)$ is the Kronecker delta function. This equation is rewritten as follows.

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(N-1) & r(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} r'(0) \\ r'(1) \\ \vdots \\ r'(N-1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (7)$$

The autocorrelation function r'(k) is obtained by resolving the above equation.

Comparing the operation in the time domain with the operation in the frequency domain from the viewpoint of computatinal complexity, the former is expensive. The computational complexity of the time domain operation is $O(N^3)$, while the computational complexity of the frequency domain operation is $O(N\log_2 N)$. What is worse that high precision is required to resolve equation (7).
Therefore, the frequency domain method was used.

B. ADAPTIVE DENSITY PULSE MODEL

The authors propose adaptive density pulses (ADP) as the synthesis filter's excitation signal. An ADP is a pulse train which is located with a constant interval, namely, with a constant density in the subframe, but the density is different subframe by subframe. The ARMA residual signal is divided into several subframe signals in a coding frame. The ADP's density is set high when the subframe's power is high, while the ADP's density is set low when the subframe's power is low. The amplitudes of the ADP are analytically determinded to minimize the perceptually weighted errors between the original and synthetic speech signals.

Let N be the coding frame length, L the subframe length, and M the number of subframes in the coding frame. The m'th subframe's excitation signal $V^{(m)}(n)$ is described as follows.

$$V^{(m)}(n) = \sum_{i=1}^{Qm} g_i^{(m)}\delta\{n-(i-1)D_m-K_m\} \qquad (8)$$

$$n=1,2,\cdots,L$$

$$1 \le K_m \le D_m$$

where D_m, Q_m, K_m, and $g_i^{(m)}$ denotes the interval, number, phase, and amplitude of the ADP in the m'th subframe, respectively. Superscript(m)denotes that the sequence with the superscript is defined only in the m'th subframe. Time (n) is defined so as to be reset to one at each subframe's beginning. An example of ADP is shown in Fig.3. The synthesized signal $Y^{(m)}(n)$ is represented by the convolution between the synthesis filter's impulse response h(n) and the excitation signal $V^{(m)}(n)$.

$$Y^{(m)}(n) = \sum_{j=1}^{L} V^{(m)}(j)h(n-j)$$

$$= \sum_{i=1}^{Qm} g_i^{(m)}h\{n-(i-1)D_m-K_m\} \qquad (9)$$

On the condition that D_m and Q_m are given, the amplitudes $g_i^{(m)}$ and phase K_m of the m'th subframe's ADP are determined so as to minimize the total squared error

$$E^{(m)} = \sum_{n=1}^{L} [S_w^{(m)}(n) - Y_w^{(m)}(n)]^2, \qquad (10)$$

where $S_w^{(m)}(n)$ denotes the weighted speech signal after subtracting the contributions carried over from the previous coding frame and subframes, and $Y_w^{(m)}(n)$ denotes the weighted synthetic signal. After differentiating E in terms of the amplitudes $g_i^{(m)}$ $(i=1,2,\cdots,Q_m)$, let them be zeros. Then the following equations hold.

$$\sum_{i=1}^{Qm} g_i^{(m)}R_{hh}(I,J) = R_{sh}^{(m)}(J) \qquad (11)$$

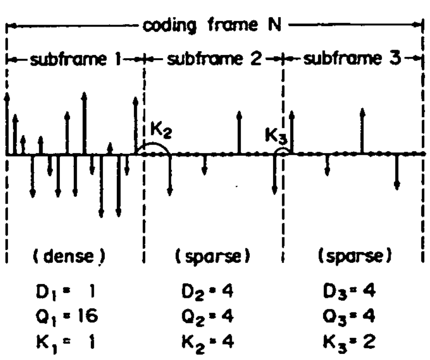

Fig.3 An example of ADP with N=48 and L=16.

$$I=(i-1)D_m + K_m,\ J=(j-1)D_m + K_m$$

$$1 \le i,\ j \le Qm$$

where $R_{hh}(I,J)$ and $R_{sh}^{(m)}(J)$ represents

$$\sum_{n=1}^{L} h_w(n-I)h_w(n-J)$$

and

$$\sum_{n=1}^{L} S_w^{(m)}(n)h_w(n-J),$$

respectively. The sequence $h_w(n)$denotes the weighted synthesis filter's impulse response. The D_m sets of candidates for the ADP's amplitudes are given by solving the equations (11) in terms of $g_i^{(m)}$ for $K_m = 1,2,\cdots,D_m$. The Dm kinds of minimum squared errors of $E^{(m)}$, $E_{min}^{(m)}(K_m)$, are represented by substituting Eq.(11) into Eq.(10).

$$E_{min}^{(m)}(k_m) = \sum_{n=1}^{L} [S_w^{(m)}(n)]^2 - \sum_{i=1}^{Qm} g_i^{(m)}R_{sh}^{(m)}(I)$$

$$I = (i-1)D_m + K_m \qquad (12)$$

Therefore the optimum phase of the ADP can be set to such a K_m that maximizes the second summation on the right side of Eq.(12). The optimum amplitudes are chosen among the D_m sets of candidates from the optimum phase K_m. The perceptual weighting filter is represented by the ARMA synthesis filter H(Z) in the z transform domain by

$$W(Z) = H(rZ)/H(Z) \qquad (13)$$

where $0 \le r \le 1$.

3. EXPERIMENTAL RESULTS

The inverse of the spectral flatness measure (sfm) [7] of the ARMA residual signals was calculated to evaluate the proposed ARMA analysis by computer simulations. The inverse of the sfm is a measure of wavefrom predictability [7]. Table 1 shows a segmental sfm^{-1} for the conventional ARMA analysis method (Method 0) and the proposed two methods, which are methods with means for eliminating the fine harmonic structure in the frequency domain (Method 1) and in the time domain (Method 2). The $sfm^{-1}seg$ is an averaged inverse of the sfm in dB units. Speech samples used for the simulations were one short Japanese sentence utterd by two male and two female speakers. The analysis frame length, the AR order, and the MA order, was 256 samples, 8, and 4, respectively. From Table 1, it can be seen that the conventional method was improved by the proposed methods.

Table 1 Sfm^{-1}_{seg} of the ARMA residnal signal.

	Male 1	Male 2	Female 1	Female 2
Method 0	8.64	5.65	9.77	8.14
Method 1	5.07	4.08	5.97	4.79
Method 2	5.33	3.96	6.14	5.00

Computer simulations were conducted to evaluate the proposed ARMA-ADP coder's performance. Table 2 shows the coding parameters and the bit allocation for them. The coding frame length and subframe length were set to 240 and 40 samples at an 8 kHz sampling rate, respectively. Two kinds of densities for ADP were used, that is, the pulse interval was two samples for dense ADPs and four samples for sparse ADPs. The number of subframes where ADP was dense was two per coding frame. The AR parameters were quantized after being converted to a log-area ratio. The MA parameters were directly quantized. ADP was quantized by a vector quantizer (VQ). The codebook was designed with 30000 training vectors generated from real speech samples using the LBG algorithm. The speech samples used for subjective and objective tests were the same ones mentioned above. They were different from the speech samples used for the VQ's design.

Table 3 shows the segmental SNR for the proposed coder at 8 kbps. The synthetic speeches had high segmental SNRs, and they were high-quality and comparable with 6-bit log PCM.

Table 2 Coding parameters.

Bit rate	8 kbps
Sampling rate	8 kHz
Frame length	
Analysis	32 ms.(256 samples)
Coding	30 ms (240 samples)
Pitch analysis	1 st order
Bit allocation	12 bits
ARMA analysis	AR 8th , MA 4th
Bit allocation	48 bits
ADP	
Subframe length	40 samples
Pulse interval	2 sample (2 subframe)
	4 samples(4 subframes)
VQ Dimension	5
Size	10 bits
Bit allocation	180 bits

4. CONCLUSION

The authors have proposed an ARMA based speech coding method with a new excitation signal model. An improved ARMA analysis has been developed. This analysis is based on LPC analysis and spectral inversion, in other words, zero to pole conversion, hence, it requires less computation, since the FFT algorithm and LPC analysis can be directly utilized. The MA parameter's misestimations are avoided by eliminating the spectral harmonic structure within the AR residual signal. The excitation signal is modeled as a pulse train whose density is varied subframe by subframe depending on the residual signal's power. The amplitudes of the pulse train are analytically solved to minimize the perceptually weighted errors between the original and synthetic signals. The subjective speech quality of the coder was comparable to that of 6-bit log PCM from an informal listening test's result.

Table 3 Coder's performance.

	Male 1	Male 2	Female 1	Female 3
SNRseg	15.6	14.3	15.7	15.3

REFERENCES

[1] K.Irie, Y.Tada and K.Honma, "APC-AB Codec Modules Operating at 16 and 8 kbit/s," IEEE J-SAC, vol.6, No.2, Feb. 1988.
[2] Y.Yatsuzuka, S.Iizuka and T.Yamazaki, "A Variable Rate Coding by APC with Maximum Likelihood Quantization from 4.8 kbit/s to 16 kbit/s," Proc. ICASSP, pp.3071-3074, April 1986.
[3] B.S.Atal and J.R.Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates," Proc. ICASSP, pp.614-617, April 1982.
[4] K.Ozawa and T.Araseki, "High Quality Multi Pulse Speech Coder with Pitch Prediction," Proc. ICASSP, pp. 1689-1692, April 1986.
[5] M.R.Schroeder and B.S.Atal, "Code Excited Linear Prediction(CELP):High Quality Speech at Very Low Bit Rates," Proc. ICASSP, pp.937-940, April 1985.
[6] S.Ishizaki, "Analysis of Speech Based on Stochastic Process Model," Bulletin of the Electrotechnical Laboratory, 39, 12, pp.881-902, 1975.
[7] N.S.Jayant and P.Noll, Digital Coding of waveforms, Englewood Cliffs, NJ: Prentice Hall, 1984.

Organization ____________ Bldg./Room PK2

U. S. DEPARTMENT OF COMMERCE
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450
IF UNDELIVERABLE RETURN IN TEN DAYS

OFFICIAL BUSINESS

AN EQUAL OPPORTUNITY EMPLOYER

2005

Technology Center